FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

Candidate:
Soodeh Farokhi, adaptive distributed systems doctoral college

Reviewers:
Priv.-Doz. Dr. Ivona Brandic, Vienna University of Technology
Uni-Prof. Dr. Erich Schikuta, University of Vienna

Title of the dissertation:
Quality of Service Control Mechanisms in Cloud Computing Environments

Abstract:
The growth in popularity of the Internet, along with the rapid development of processing and storage technologies, has brought a paradigm shift in the way computing resources are provisioned. The technological trend today is to offer computing resources as services, leased and exposed via the Internet in a pay-as-you-go and on-demand fashion, called cloud computing. The interest in cloud computing is growing in both industry and academia, so the number of cloud providers offering their services, and the number of cloud customers interested in using such services is rapidly increasing.

Cloud infrastructure providers are trying to reduce their operating costs while offering their services with higher quality; something they strive to do to stand out among other providers. However, this is becoming challenging as providing such services needs operating large-scale and geographically distributed data centers. On the other hand, the main purpose of customers in using clouds is to achieve a high quality of service (QoS) while reducing their overall costs. Given the variety of offered services in terms of quality and cost, customers are encouraged to simultaneously use services from multiple cloud providers, known as multi-cloud. However, utilizing multi-cloud brings a new set of open challenges, such as selecting and composing the most appropriate services. Furthermore, despite the critical need of customers in having predictable service performance, in general cloud providers do not yet offer any performance guarantees. This gap is due to the complexity of practically addressing this issue in a cost-effective way. Such a complexity mainly comes from the dynamic nature of the cloud, unpredictable workloads, and non-linearity of mapping performance measurements into required cloud resources. Hence, controlling the trade-off between QoS and cost is a challenging goal for both cloud infrastructure providers and customers.

This thesis investigates models, algorithms, and mechanisms to tackle this trade-off from both perspectives. More specifically, in the scope of this thesis, we first take the cloud provider viewpoint by proposing an approach for virtual machine placement across geographically distributed infrastructures. In this approach, a Bayesian network model is used to address decision making under uncertainty. Then, we address the trade-off between QoS and cost from the cloud customer point of view by facilitating the utilization of the multi-cloud paradigm. We propose a service selection approach using prospect theory to rank the comparable service offerings. Furthermore, to guarantee the performance objectives of customers, we propose autonomic resource provisioning techniques. To this aim, control theory is used to design resource provisioning controllers, and fuzzy control is utilized to coordinate multiple controllers toward meeting the service performance objectives in a cost-effective manner. Finally, the evaluations of these contributions in comparison with the state-of-the-art approaches are presented.